The gender gap in mathematics achievement: Evidence from Italian data.

Dalit Contini[1], Maria Laura Di Tommaso [2], Silvia Mendolia[3]

Abstract

Gender differences in the STEM (Science Technology Engineering and Mathematics) disciplines are widespread in most OECD countries and mathematics is the only subject where typically girls tend to underperform with respect to boys. This paper describes the gender gap in math test scores in Italy, one of the countries displaying the largest differential between boys and girls according to PISA, using the data from an Italian national level learning assessment, involving children in selected grades from second to tenth. We first analyse the magnitude of the gender gap using OLS regression and school fixed-effect models for each grade separately. Our results show that girls systematically underperform boys, even after controlling for an array of individual and family background characteristics including math self-beliefs, and that the average gap increases with children's age. We then study the gender gap throughout the test scores distribution, using quantile regressions and metric-free methods, and find that the differential is small at the lowest percentiles, but large among top performing children. Finally, we estimate dynamic models relating math performance at two consecutive assessments. Lacking longitudinal data, we use a pseudo panel technique and find that girls' average test scores are consistently lower than those of boys at all school years, even conditional on previous scores, meaning that gender-specific effects in favour of boys operate at all stages of schooling.

JEL: J16; I24; C31.

Keywords: Math gender gap, education, school achievement, inequalities, cross-sectional data, pseudo panel estimation, quantile regression

---

[1] Dept of Economics and Statistics Cognetti de Martiis, Lungo Dora Siena 100, Torino, Italy. dalit.contini@unito.it

[2] Dept of Economics and Statistics Cognetti de Martiis, Lungo Dora Siena 100, Torino, Italy. marialaura.ditommaso@unito.it and Collegio Carlo Alberto, Moncalieri, Italy.

[3] School of Accounting, Economics, and Finance, University of Wollongong, Northfields Avenue, North Wollongong, NSW 2522. smendoli@uow.edu.au

1. Introduction

The traditional gender gap in educational outcomes advantaging boys has been completely filled up in most industrialized countries, and has now reversed in favour of girls. Girls tend to do better than boys in reading test scores, in grades at school, in the propensity to choose academic educational programs in upper secondary school, in tertiary education attendance and graduation rates. In this perspective, there is now an extensive literature addressing the underperformance of boys (Legewie and DiPrete 2012). However, boys keep doing better than girls in math tests. According to the PISA international assessment (OECD 2015), the average gender differential within OECD countries in mathematics at age 15 is 0.11 standard deviations in favour of males. Italy is one of the countries with the largest gap, equal to 0.24 standard deviations. The presence of a substantial females' disadvantage in math is of particular importance, because it is likely to be a cause of the critically low share of women choosing STEM (Science Technology Engineering and Mathematics) disciplines at university, of gender segregation in the labour market, and gender pay gaps (European Commission 2006, 2012, 2015; National Academy of Science, 2007).

Several explanations have been proposed for the existence of the gender gap in mathematics. Some scholars refer to biological factors (Baron-Cohen and Wheelwright 2004, Baron Cohen et al 2001). However, as shown by international assessments (OECD 2015, Mullis et al, 2012) the gender gap in math differs substantially across countries. Hence, "nature" cannot be the only account for the females' disadvantage in math; there must be alternative explanations related to societal and cultural factors, supporting the existence of "nurture" effects. In this perspective, some scholars (Guiso et al., 2008, de San Roman and de La Rica Goiricelaya, 2012; OECD 2015) show that the gender gap in math in the PISA survey is negatively related to country level indexes of gender equality.

Focusing on micro-level mechanisms, the literature emphasizes the importance of parents and teachers' beliefs about boys and girls capacities. Sex-stereotypes in parent's evaluation of children abilities have been found to affect achievements and children's self-perception (amongst others: Jacobs 1991; Jacobs and Bleeker 2004; Jacobs and Eccles 1992; Bhanot and Jovanovic 2009; Twenge and Campbell 2001). Fryer and Levitt (2010) show that parental expectations regarding math are lower for girls than boys even after accounting for test scores. However, Cornwell et al (2012) and Robinson et al (2014) find that teachers rated girls' math skills higher than those of observationally similar boys.

Whatever the reason (the causal direction is difficult to assess), girls display less math self-efficacy (self-confidence in solving math related problems) and math self-concept (students' beliefs in their own abilities), and more anxiety and stress in doing math related activities (OECD 2015, Heckman and Kautz 2012, 2014; Lubienski et al 2013, Twenge and Campbell 2001). As demonstrated by the recent work by Heckman and colleagues (e.g. Heckman and Kautz 2012, 2014; Heckman and Mosso, 2014), non-cognitive abilities including motivation and self-esteem are important predictors of success in life and in the labour market. In this perspective, the females' lower self-esteem in math could be responsible of their relatively poorer performance in STEM subjects and future educational choices and occupational outcomes.

Socioeconomic status, parental education and occupation indubitably play a major role on school achievement, including math performance. In particular, parenting styles and familial activities – typically related to social background – such as involvement in children's homework, are relevant (Bhanot and Jovanovic 2009, Del Boca, et al 2014, Brilli et al. 2016). Home experience differ in stereotypical gendered ways, and according to Lubienski et al (2013), somewhat unexpectedly, more strongly among children of high socioeconomic status. Whether these factors also contribute to determining the gender gap in math test scores is still unclear. However, there is empirical evidence that girls with mothers working in math-related occupations lag behind boys as much as those whose mothers are not in math-related occupations (Fryer and Levitt, 2010; OECD 2015)

Schools and educational methods and practices also seem to matter. The educational literature provides evidence that different approaches to math and physics can decrease gender inequality in achievements. Problem solving, class-discussions and investigative work and cognitive activation strategies have been found to improve girls' performances (Boaler 2002; Zohar and Sela 2003; OECD 2015). In addition, Boaler et al. (2011) and Good, Woodzicka, and Wingfield (2010) analyse other factors affecting lower achievement in math and science related to sex-stereotypes, such as images shown in textbooks and show that girls' proficiency increases by using counter-stereotypic pictures with female scientists.

From a policy perspective, it is important to describe when the gap first shows up. Does it already exist at the beginning of primary school? The existing evidence is mainly based on the US dataset "Early Childhood Longitudinal Study, Kindergarten Class of 1998– 1999" (ECLS-K) following students from kindergarten through eighth grade. The main finding deriving from these data is that the math gender gap starts as early as in kindergarten and increases with the age of the child (Robinson and Lubiensky, 2011; Fryer and Levitt, 2010; Penner and Paret, 2008). Another relevant result is that the math gender gap becomes higher for top performing students. Initially boys appear to do better than girls among well performers and worse at the bottom of the distribution; by third grade, the gender gap, still larger at the top, appears throughout the distribution. Moreover, the male advantage among high performers is largest among families with high parental education. Girls appear to lose ground in math over time in every family structure, racial group, and level of the socio-economic distribution (Fryer and Levitt, 2010). These findings are confirmed by PISA international data, on children of age 15 (OECD, 2015).

There is also some evidence on UK… Instead, evidence on other European countries is scant (Camilla su DEU?), and there are currently no contributions on the achievement gender differential in Italy.

This paper contributes to the existing literature in a number of ways. Firstly, it provides detailed descriptive evidence on the gender gap in math test scores in Italy, one of the countries with the largest differential favouring boys over girls at age 15 (OECD 2015). We exploit the cross-sectional data of the National Assessment carried out by INVALSI for year 2013, testing the entire population of Italian children in school years 2, 5, 6, 8 and 10.[4] In particular, we analyse the gender gap at different school years to identify when the gap first emerges, and to study its evolution throughout compulsory schooling. Secondly, by focusing on differentials along the entire test score distribution, we apply quantile regression and metric free methods to analyse where the girls' disadvantage is more severe. Lastly, we estimate dynamic models relating math performance at two consecutive assessments. Lacking longitudinal data, we use a pseudo panel technique that allows identifying "new" gender effect operating between the two surveys and disentangling them from carryover effects of previously established inequalities. We find that performance of girls is always lower than that of boys, even given prior test scores. Altogether, this body of evidence shows that despite the magnitude of the gap differs, the qualitative findings on gender inequalities in math test scores observed in US data also apply to Italy.

[4] INVALSI stands for "Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione" (National Institute for the evaluation of education and training).

2. Modeling strategies

Cross-sectional models

Since test scores are not measured on the same scale at different school years, the gender gap on original scores is not comparable across grades. For this reason, we use standardized scores and the gender gap results show by how many standard deviations girls and boys differ.

First, we focus on the total effect of gender on average math achievement. To this aim, we run the basic OLS model with standardized test scores as dependent variables, gender as the independent variable of interest, and a set of control variables describing maternal and parental education, socio-economic status of the family, geographical area and number of siblings.

Yet, if school characteristics influence children' learning, the effect of gender might operate both indirectly via school choices and directly net of school characteristics. The existence on an indirect effect could play a role in particular after tracking into different educational programs has taken place (see section 3), but in principle it may apply even at earlier stages of schooling. In fact, students attending the same school are exposed to a similar environment in the student body composition (in terms of gender, socioeconomic status, immigrant background), learning targets, educational practices, gender stereotypes, that might affect the performance of girls and boys differently. In this perspective, we estimate the direct effect of gender on math achievement with school fixed effects linear models that exploiting only within-school variability, deliver valid estimates of the gender gap given individual controls and (observed and unobserved) school characteristics.

Third, we shift the focus from the expected value of test scores to the entire test score distributions of girls and boys and analyse gender differentials at different points of the ability distribution. To this aim, we estimate quantile regression models (Koenker and Basset, 1978). In essence, with these models we inspect the gender gap at different percentiles of the distribution, and assess whether female's disadvantage in math exists throughout the distribution, or instead if it is stronger among low performing or top performing children. In the simplest case where gender is the only explanatory variables, the quantile regression coefficient gives the difference between the score corresponding to a specific percentile of the girls' distribution and the score corresponding to the same percentile of the boys' distribution.

A limitation of the methods described so far is that test scores are considered on the interval scale, implicitly relying on stringent psychometric assumptions. For example, that there is the same difference in cognitive ability between two children scoring 0.70 and 0.80 and between two children scoring 0.40 and 0.50. An alternative approach is given by metric-free methods that rely instead on the relative position that girls and boys occupy in the overall ranking. Following Robinson and Lubienski (2011), we analyze the gender gap throughout the distribution by estimating at specific percentiles θ the following:

$$\lambda_\theta = \begin{cases} \frac{\varphi_M(\theta)}{\varphi_M(\theta)+\varphi_F(\theta)} & if\ \theta < 50 \\ \frac{1-\varphi_F(\theta)}{2-(\varphi_M(\theta)+\varphi_F(\theta))} & if\ \theta \geq 50 \end{cases} \tag{1}$$

where $\varphi_M(.)$ and $\varphi_F(.)$ are the cumulative distribution functions of males and females at the θth percentile of the overall distribution. Values of λ_θ below 0.5 indicate a girls' disadvantage. For example, $\varphi_F(20)$ is the percentage of females below or at the 20^{th} percentile of the overall distribution. If $\varphi_F(20) > \varphi_M(20)$, more girls perform below the 20^{th} percentile than boys and thus λ_θ<0.50. Instead, $1 - \varphi_F(80)$ is the percentage of females above or at the 80^{th} percentile of the overall distribution. So, if $1 - \varphi_F(80) < 1 - \varphi_M(80)$, a lower share of girls perform above the 80^{th} percentile as compared to the share of boys, and, again, λ_θ<0.50.

Dynamic models

Cross-sectional analyses do not allow exploring the mechanisms underlying the *development* of inequalities as children grow. In fact, cross-sectional regression coefficients at age *t* represent the effects accumulated up to age *t* and do not allow distinguishing between new effects operating between two successive assessments and carryover effects of preexisting achievement gaps between girls and boys. Moreover, coefficients based on standardized test scores also depend on the achievement variability at each assessment. Hence, if this variability increases between two surveys for reasons not related to gender (for example, due to increasing differences across socioeconomic levels), we might observe a diminishing gender gap even if there are no forces at work making girls and boys move closer (Contini and Grand, 2016).

In this perspective, we aim at estimating a simple dynamic model, relating achievement at a given time point (*t*=2) to previous achievement (at *t*=1) and individual characteristics, including gender. In the absence of longitudinal data, we use pseudo-panel techniques proposed by De Simone (2013) and Contini and Grand (2016), allowing to estimate simple dynamic models with repeated cross-sectional data. The basic idea is that the unobserved lagged dependent variable can be replaced by a predicted value from an auxiliary regression using individuals observed in previous cross-sections. This strategy delivers consistent estimates under quite restrictive conditions – for example, if there are no time-varying exogenous variables or the time-varying exogenous variables are not auto-correlated (Verbeek and Vella, 2005). These conditions apply to our case study, because the explanatory variable of main interest is gender and the other control variables are time-invariant sociodemographic variables.[5]

In this paper, we apply the method adopted in Contini and Grand (2016). To illustrate its rationale, consider two cross sectional assessments using a single scale to measure achievement (i.e. "vertically equated" scores). Subsequent scores follow the relation: $y_{i2} = y_{i1} + \delta_i$, where δ_i is achievement growth, that may vary across individuals and depend linearly on individual characteristics x_i and previous achievement: $\delta_i = \Delta + \beta x_i + \theta y_{i1} + \varepsilon_{i2}$. Under these assumptions, the dynamic model relating achievement at the two occasions is $y_{i2} = \Delta + (1+\theta) y_{i1} + \beta x_i + \varepsilon_{i2}$. The parameter of interest is β, representing the difference between test scores at t=2 of a boy and a girl with identical performance at t=1. Hence, β captures gender inequalities developed between the two surveys. Instead, θ are carry-over effects of inequalities already existing at t=1. When achievement scores are not equated, the relation between subsequent scores is: $y_{i2} = \tilde{y}_{i1} + \delta_i$, where $\tilde{y}_{i1}$ represents achievement at t=1 in the measurement scale employed at t=2. Assuming that $\tilde{y}_{i1} = \varphi + \omega y_{i1}$ (where φ and ω are not known and not identifiable), the dynamic model becomes:

$$y_{i2} = \varphi(1+\theta) + \Delta + \omega(1+\theta) y_{i1} + \beta x_i + \varepsilon_{i2} \quad (2)$$

If test scores are measured on different scales, θ is always unidentified. Instead, β is identified and can be estimated even with repeated cross-sectional data.

In the first step, we estimate the cross sectional model for test scores at t=1: $y_{i1} = \mu_1 + \rho x_i + \delta w_i + \varepsilon_{1i}$, where w is an appropriate instrumental variable affecting achievement at t=1 but not affecting achievement at t=2 given achievement at t=1. Following Contini and Grand (2016), we use the month of birth, since there is widespread evidence – confirmed by our data – that younger children generally underperform their older peers, in particular at early school stages. Further, the identifying assumption is that achievement at a given school year is not affected by the age of the child given earlier achievement.[6]

[5] Notice that the inclusion of school characteristics in the model would invalidate the estimation,because school features are typically correlated to the error term (that incorporates innate ability), because higher ability children usually choose schools with more favorable characteristics (Contini and Grand, 2016). Similar conclusion would apply if we were to include other endogenous variables capturing behavior and attitudes.

[6] The use of the season of birth as an instrumental variable to account for the endogeneity of children's age on later outcomes has recently been questioned by Buckles and Hungerman (2013). They argue that, contrary to common belief, the season of birth is not totally idiosyncratic; in fact in USA winter births are disproportionally represented by teenagers and unmarried mothers. Yet, since we make a different use of the month of birth, we believe this criticism does not apply to our paper.

In the second step, we substitute y_1 with its OLS estimate $\hat{y}_1$ and plug it in model (2). This introduces measurement error $\hat{y}_1 - y_1$ in previous scores; however, due to properties of OLS estimates, this measurement error (which enters the error term) will be uncorrelated to x and $\hat{y}_1$. Hence, standard estimation of model $y_{i2} = \mu_2 + \gamma \hat{y}_{1i} + \beta x_i + u_{2i}$ will deliver consistent estimates of β. Clearly, the drawback is that the standard errors will be largely inflated, so for reliable estimation large samples and a good instrument are needed.

3. Italian Education system and data

The Italian education system is organised in three stages. Students attend primary school from the age of 6 until the age of 11 years old. At the end of primary school, they enrol in middle school, and remain within the same institution from the age of 11 until the age of 14 years old. High school begins at the age of 14 and lasts five years, but compulsory education is up to 16 years old, so a relevant share of children does not attain the upper secondary school diploma. At the end of middle school, students choose among different kinds of high schools, with significant differences in the curriculum. These educational programs are broadly classified into three main types: the Lyceum, the Technical High School and the Vocational High School. The curriculum is generally organised at national level and all high schools have to offer some compulsory subjects (Italian, Mathematics, Sciences, History, one or two foreign languages and Physical Education). However, there are significant differences in terms of the time allocated to each subject, and the specialised field of studies. Lyceums generally provide a higher level academic education, with a specialisation in the humanities, sciences, languages or arts. Technical institutes usually provide students with both a general education and a qualified technical specialization in a particular field (e.g.: business, accountancy, tourism, technology). Vocational institutes have specified structures for technical activities, with the objective of preparing students to enter the workforce.

This study uses data from the National Test INVALSI for 2013, aiming at assessing the reading and mathematical skills of Italian pupils. Since 2009, all Italian children have been tested by the Italian Institute for the Evaluation of the Education System (INVALSI) during school years 2, 5, 6, 8 and 10. More than half a million students in each grade sit this test each year.

INVALSI assesses the overall population of students enrolled in Italian schools but a subsample of schools and students takes the tests under the supervision of an external inspector. In our analysis, we only use the subsample of children whose test was supervised by an external inspector. We also restrict the sample to native children, mostly because recent migrants may be enrolled in classes that are not necessarily aligned with their age, depending on their level of fluency in Italian. Further, immigrants experience grade repetition more frequently than native students. Our final sample includes around 23,000 observations from year 2; 22,000 from year 5; 24,000 from year 6 (first year middle school); 25,111 from year 8 (third year of middle school) and 34,000 from year 10 (second year high school).

In addition to test scores, INVALSI data includes information on parental characteristics and family background, collected from a questionnaire submitted to the students and from school board records. Among the variables on family background, for years 5, 6, and 10, INVALSI provides a synthetic indicator of economic and socio-cultural status (ESCS) similar to that available in PISA, by taking into consideration parental educational background, their employment and occupation, and home possessions. Moreover, children in years 5, 6 and 10 are asked questions regarding their beliefs on their own abilities in the subjects of the test, and on the importance of language and math for their future studies, life, and career. Full descriptive statistics for the variables used in the estimation are provided in tables A1 and A2.

4. Results

Table 1 shows average test scores in mathematics for girls and boys, by school year and gender, and the standardized gender gap. The gender gap in math already exists at grade 2, and increases from nearly 10% of a standard deviation to almost 30% in grade 10, with a slight decrease between years 5 and 6.

Table 1 – Average test scores in Math and the gender gap

Original scores (% of correct answers)	Year 2	Year 5	Year 6	Year 8	Year 10
All	54.9 (20.7)	55.6 (18.8)	45.3 (16.7)	51.8 (18.9)	42.7 (17.8)
Boys	55.9 (21.1)	57.4 (19.0)	46.8 (17.3)	53.8 (19.0)	45.2 (18.6)
Girls	53.9 (20.2)	53.8 (18.4)	43.8 (16.0)	49.7 (18.5)	40.2 (16.6)
Standardized gender gap	0.0907	0.202	0.170	0.206	0.286

Note: Standard deviation in brackets. All gender gaps are significant at the .001 level.

<u>Average gender gap. OLS and school fixed-effect models</u>

Table 2 shows the results for the gender variable in OLS and school fixed effect regression models with region of residence, maternal and paternal education, ESCS and number of siblings as controls. The full set of results (also including the interaction between gender and ESCS) is shown in the Appendix.

Table 2 – The gender gap (girls-boys) in math test scores. OLS and school fixed effects.

	Year 2	Year 5	Year 6	Year 8	Year 10
OLS	-0.105	-0.183	-0.168	-0.184	-0.304
	(0.014)***	(0.014)***	(0.013)***	(0.012)***	(0.010)***
Fixed Effects	-0.099	-0.185	-0.168	-0.191	-0.281
	(0.013)***	(0.013)***	(0.013)***	(0.011)***	(0.009)***
N					
R^2					

Note. Std errors are in brackets. * indicates that the underlying coefficient is significant at 10% level, ** at 5% and ***1%. All models include region of residence, maternal and paternal education, ESCS (not available in years 2 and 8) and number of siblings (not available in years 2 and 8). They also include an interaction effect between gender and ESCS. The results reported in this table refer to the gap at ESCS=0 (approximately the mean value of ESCS).

Gender has a significant and sizable effect on test scores in mathematics at all ages. Since gender is nearly independent of the included controls, we do observe major changes in the OLS estimates as compared to the raw gap. There are little changes also when adopting a school-fixed effect model, meaning that there is no substantial indirect effect of gender via school characteristics, not even at year 10, where schools differ markedly and the choice between school types is strongly related to individual and family characteristics.

The complete set of results on the effects of all control variables are reported in the Appendix. Parental education and socio-economic status are strong determinants of students' achievement. Consistently with the literature, we find that girls benefit less from family resources than boys do, as shown by the negative sign (although not always significant) of the interaction term between gender and ESCS. In addition, we observe dramatic differences across geographical areas and differences between children with different number of siblings. Interestingly, while the gender gap changes little with school-fixed effect estimation, in year 10 within-school differences across social groups (identified by parental education and ESCS) are relatively weak as compared to OLS estimates. Hence, the sorting into different high school types – a highly gendered and socially selective process – does not appear to play a relevant mediating role on the establishment of gender differences in math achievement, while it has a strong mediating effect on socioeconomic background achievement inequalities.

Quantile regression and metric-free analysis of the test score distributions

Table 3 summarizes the gender effect from quantile regression. These findings show that the gap between girls' and boys' performance in mathematics – always non-negative – increases throughout the distribution in all school years. In year 2, we observe no differences at the tenth percentile of the girls and boys' distributions; the gap between girls and boys at the first quartile of the distribution is about 0.05 standard deviations, but it is more than 0.14 standard deviations at the third quartile. These gaps widen in later grades. By year 10, girls in the bottom quartile underperform boys by nearly 0.3 standard deviations, whereas the gap between students in the top 10% of the distribution is almost 0.5 standard deviations.[7] Our results confirm the finding for the US that the girls disadvantage increases over time and is larger at the top of the distribution (Robinson and Lubiensky 2011; Fryer and Levitt 2010).

Table 5 - Gender gap in achievements in Mathematics – Quantile Regression

	Year 2	Year 5	Year 6	Year 8	Year 10
Q10	0.000	-0.132	-0.073	-0.117	-0.131
	(0.008)	(0.018)***	(0.014)***	(0.037)***	(0.011)***
Q25	-0.048	-0.183	-0.127	-0.233	-0.180
	(0.023)***	(0.018)***	(0.015)***	(0.028)***	(0.012)***
Q50	-0.145	-0.203	-0.181	-0.233	-0.268
	(0.038)***	(0.027)***	(0.013)***	(0.034)***	(0.013)***
Q75	-0.145	-0.244	-0.256	-0.233	-0.341
	(0.021)***	(0.023)***	(0.021)***	(0.032)***	(0.015)***
Q90	-0.145	-0.201	-0.282	-0.233	-0.423
	(0.026)***	(0.024)***	(0.026)***	(0.031)***	(0.025)***

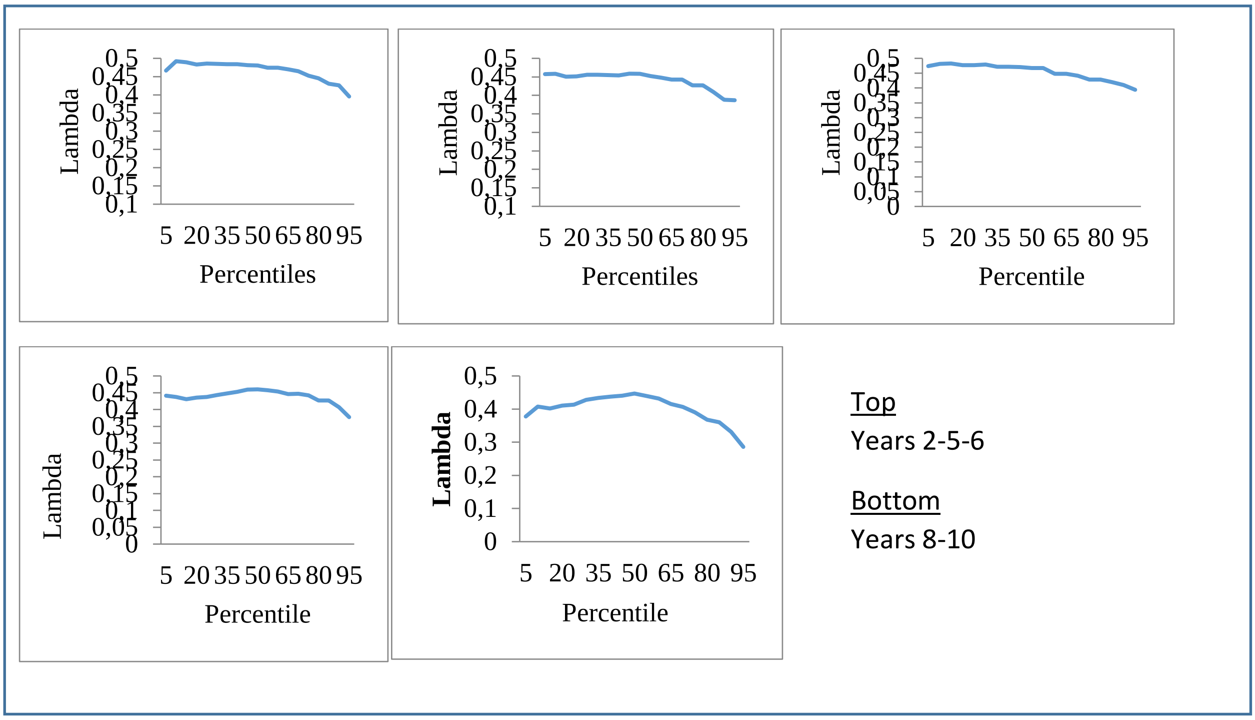
Note: Std errors are in brackets. * indicates that the underlying coefficient is significant at 10% level, ** at 5% and ***1%. All models include region of residence, maternal and paternal education, ESCS (not available in years e and 8 data) and number of siblings (not available in years e and 8 data).

In order to check the robustness of our estimates in the quantile regression, we compute the metric-free gap at different points in the achievement distribution (see Section 2). Metric-free findings qualitatively confirm the results obtained with quantile regression. Figure 1 presents metric-free measures of the math gap throughout the test scores distribution in year 2, 5, 6, 8 and 10. Values of λ below 0.5 indicate a boys' advantage, while above 0.5 indicate a girl's advantage. Figure 1 shows that the gap significantly favours males at all percentiles. For instance, in year 2, λ_{95} is equal to 0.4, which means that if girls and boys where equal in number, the top 5% of the distribution would be composed by 40% of girls and 60% of boys. On the other hand, the proportion of boys and girls is nearly even (λ almost equal to 0.5) at the 10th percentile of the distribution. Consistently with the quantile regression findings, the largest

[7] OECD 2015 reports similar results for Italian children in PISA data at age 15.

gender gap is observed in year 10 at the top 10% of the grade distribution, where the adjusted proportion of females is as low as 33%.

Figure 1 – Metric-free gender gap in achievements in Maths throughout the distribution



Note: λ equal to 0.5 means that boys' and girls' grades are aligned. λ values closer to 0 benefit boys while values closer to 1 favour girls.

Dynamic model estimation

Table 4 presents results for the gender dummies from the pseudo-panel methodology[8]. In this framework, the coefficients measure the extent to which achievement growth between t=1 and t=2 differs across gender, when comparing two children performing at the same level in t=1. Columns 1, 2, 4, 6 and 8 include results from cross-section models while the other columns report results for dynamic models. Results from pseudo-panel modelling show that when comparing girls and boys with the same background characteristics and prior scores, at the following assessment on average the boys will score better than the girls. For instance, according to our estimates, between grades 2 and 5, if we take girl and a boy with the same performance at grade 2, the expected score at grade 5 for the girl will be 0.113 standard deviations below the corresponding value for the boy at grade 5. The progressive deterioration of female's performance relative to males is observed between all pairs of assessments, including between year 5 and 6, where at the cross sectional level we observed instead a slight

[8] Full estimates available from the authors upon request.

reduction of the gender gap. This result may appear awkward at first sight; we observe a decreasing standardized gender gap despite the existence of mechanisms making girls keep losing ground relative to boys if other inequalities operate (for example, across social backgrounds) that make the achievement variability increase substantially between the two assessments (see Contini and Grand, 2016, pg…).

Table 6 – New gender effects in Mathematics - Pseudo panel estimation (rivedi valori)

	Y-2 CS	Y-5 CS	Y-5 Dyn	Y-6 CS	Y-6 Dyn	Y-8 CS	Y-8 Dyn	Y-10 CS	Y-10 Dyn	Y-10 Dyn (base y6)
Female	-0.105 (0.014) ***	-0.183 (0.014) ***	-0.115 (0.017) ***	-0.172 (0.014) ***	-0.043 (0.024) *	-0.219 (0.013) ***	-0.170 (0.026) ***	-0.342 (0.012) ***	-0.251 (0.092) ***	-0.322[9] (0.023) ***
Month of birth	-0.032 (0.022) ***	-0.021 (0.002) ***		-0.015 (0.002) ***		-0.004 (0.002) ***		-0.002 (0.016)		

Note: Std errors are in brackets. * indicates that the underlying coefficient is significant at 10% level, ** at 5% and ***1%. All models include region of residence, maternal and paternal education, ESCS (not available in years 2 and 8) and number of siblings (not available in years 2 and 8).

[9] Nota su limiti della stima dyn 10…

5. Conclusions

The paper utilises several techniques (OLS, School fixed effects, Pseudo-panel, quantile regressions, and metric free measures) to explore the gender gap in math in Italy. In 2013, Invalsi data show that boys outperform girls in math from age 7 until age 15. Results show that gender dummy for girls is negative even after controlling for many covariates related to the family socioeconomic status, geographical areas, parental education, maternal employment, preschool attendance, number of siblings'.

Pseudo panel estimations confirm that the gap is increasing with age of the child while quantile regressions show that the gender gap in math is higher for top performer kids. Metric free results confirm the quantile regression results.

Obviously, the lack of longitudinal data for Italy is a major problem for analysing changes in gender gaps across years. Unfortunately, while the improvement of the educational system seems to have been a priority of all Italian governments in the last ten years, there has been no discussion about the importance of having reliable longitudinal data to study inequalities (not only gender inequalities) in the Italian educational system.

References

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord.*; 31(1):5-17.

Baron-Cohen S, Wheelwright S. (2004) The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J Autism Dev Disord;* 34(2):163-75.

Bhanot, R. T., & Jovanovic, J. (2009). The links between parent behaviors and boys' and girls' science achievement beliefs. *Applied Developmental Science*, 13 (1), 42-59.

Boaler, J. (2002). Paying the price for 'sugar and spice': shifting the analytical lens in equity research. *Mathematical Thinking and Learning*, 4(2&3), 127-144.

Boaler, J, Altendorff, L. & Kent, G. (2011). Mathematics and science inequalities in the United Kingdom: when elitism, sexism and culture collide. *Oxford Review of Education*, 37(4), 457-484.

Brilli Y,. Del Boca D., Pronzato C. (2016) Does child care availability play a role in maternal employment and children's development? Evidence from Italy. Review of Economics of the Household,14 (1), 83-102

Brunello, G., & Checchi, D. (2005). School quality and family background in Italy. *Economics of Education Review*, 24(5), 563–577.

Contini D. & Grand E. (2015). On Estimating Achievement Dynamic Models from Repeated Cross-Sections. *Sociological Methods and Research*, doi: 10.1177/0049124115613773.

Del Boca, D., Flinn, C. & Wiswall, M. (2014). Household Choices and Child Development. *Review of Economic Studies* 81 (1), 137-185.

De San Román, A. G, & de La Rica Goiricelaya, S. (2012). Gender gaps in PISA test scores: the impact of social norms and the mother's transmission of role attitudes. *IZA Discussion Paper* 6338.'

De Simone, G. (2013). Gender unto primary the things which are primary's: Inherited and fresh learning divides in Italian lower secondary education. *Economics of Education Review*, 35(C), 12-23.

European Commission 2015. Science is a girls' thing, Newsletter Nov. 2015 http://science-girl-thing.eu/newsletters/november-2015.html

European Commission 2006, Women in Science and Technology- the Business Perspective, Luxembourg: Office for Official Publication of the European Communities

European Commission 2012, Enhancing excellence, gender equality and efficiency in research and innovation, Luxembourg: Office for Official Publication of the European Communities

Fryer, Roland G., and Steven D. Levitt. 2010. "An Empirical Analysis of the Gender Gap in Mathematics." *American Economic Journal: Applied Economics*, 2(2): 210-40.

Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *Journal of Social Psychology*, 150 (2), 132-147.

Guiso L., Monte F., Sapienza P., & Zingales L. (2008). Culture, gender, and math. *Science*, 320 (5880), 1164–1165.

Heckman, J. J. & T. Kautz (2012). Hard evidence on soft skills. *Labour Economics*, *19*, 451-464.

Heckman, J. J. & T. Kautz (2014). Fostering and measuring skills interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), The GED Myth: Education, Achievement Tests, and the Role of Character in American Life, Chapter 9. Chicago, IL: University of Chicago Press.

Jacobs, J. E. (1991). Influence of gender stereotypes on parent and child mathematics attitudes. *Journal of Educational Psychology*, 83 (4), 518–527.

Jacobs, J. E., Lanza S., Osgood D.W., Eccles J.S., Wigfield A. (2002), Changes in children's self-competence and values: Gender and domain differences across grades 1 through 12, Child Development, Vol73, pp509-527.

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: do parents matter? *New Directions for Child and Adolescent Development,* 106, 5-21.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children`s ability perceptions. *Journal of personality and Social Psychology*, 63(6), 932-944.

Koenker , R., & Basset G. (1978) Regression Quantiles, *Econometrica*, Vol 46 (1), pp. 33-50.

Legewie, J., DiPrete, T. A. (2012). School Context and the Gender Gap in Educational Achievement, *American Sociological Review* 77 (3): 463–85.

Mendolia, S., Walker, I. (2014) The effect of personality traits on subject choice and performance in high school: Evidence from an English cohort. Economics of Education Review, 43: 47-65.

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011. International results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Academy of Science (2007) Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering (2007)

OECD (2013), PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

OECD. (2015). *The ABC of Gender Equality in Education: Aptitude, Behavior, Confidence*. PISA Paris: OECD.

Penner, A.M., Parer M., (2008) Gender differences in mathematics achievement: Exploring the early grades and the extremes Social Science Research 37 (1), 239-253

Robinson Joseph P. and Lubiensky Sarah T. (2011) The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School Examining Direct Cognitive Assessments and Teacher Ratings. *American Educational Resources Journal*. 48, 2268-302

Robinson-Cimpian JP, Lubienski ST, Ganley CM, Copur-Gencturk Y (2014) Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Development Psychology* 50(4):1262-81. doi: 10.1037/a0035073.

Twenge, J. M., & W. K. Campbell. (2001). Age and birth cohort differences in self-esteem: a cross-temporal meta-analysis. *Personality and Social Psychology Review*, 5(4), 735–748.

Verbeek, M. & Vella, F. (2005) Estimating dynamic models from repeated cross-sections, *Journal of Econometrics*, **127**, 83-102

Zohar, A., & Sela, D. (2003). Her physics, his physics: gender issues in Israeli advanced placement physics classes." *International Journal of Science Education*, 25(2), 245-268.

Appendix A

Table A1: Descriptive statistics (estimation samples from Invalsi 2013)

Gender	**Year 2**	**Year 5**	**Year 6**	**Year 8**	**Year 10**
Male	48.39	49.97	50.29	50.40	50.80
Female	51.61	50.03	49.71	49.60	48.79
Missing					0.41
ESCS index					
Mean	n.a.	0.0664	0.1033	n.a.	-0.0013
Standard deviation		1.0194	0.9842		0.9795
Region of residence					
North-West	16.67	16.26	19.29	18.47	18.78
North-East	19.90	19.83	21.04	20.41	20.75
Centre	18.05	17.32	18.06	19.01	17.62
South	25.60	26.19	23.61	23.02	24.52
Islands	19.78	20.39	18.00	19.08	18.33
Maternal education					
Degree	16.49	14.21	13.23	12.70	18.95
High school	34.01	33.73	32.21	29.64	35.20
Middle school	29.49	32.76	36.86	35.27	37.56
Missing	20.01	19.29	17.70	22.39	8.29
Paternal education					
Degree	12.27	11.49	11.17	10.90	17.56
High school	29.76	28.66	27.05	25.67	31.81
Middle school	36.54	39.49	42.80	39.94	39.79
Missing	21.43	20.36	18.98	23.49	10.84
Number of siblings	n.a.			n.a.	
0		15.19	15.40		14.74
1		54.50	56.24		55.32
2		19.65	20.68		22.06
3		4.61	4.70		4.93
>=4		2.24	2.56		2.40
Missing		3.81	0.42		0.56
Type of high school attended	n.a.	n.a.	n.a	n.a.	
Lyceum					44.57
Technical HS					21.97
Vocational HS					33.46

Table A2 – Attitudes towards maths

	(% all sample)	%Girls	%Boys
What do you think of mathematics?	**Year 5**		
I am good at maths	74.49	70.33	78.64
Maths is hard	23.26	27.31	19.20
I learn maths easily	63.30	59.14	67.47
I have fun doing maths	61.18	56.75	65.61
I'd like to do more maths a school	37.16	31.68	42.65
What do you think of mathematics?	**Year 6**		
I am good at maths			
Strongly disagree	4.10	4.90	3.36
Disagree	18.73	21.50	16.01
Agree	54.93	56.00	53.87
Strongly agree	22.03	17.50	26.52
Missing	0.21	0.17	0.25
Mathematics is hard			
Strongly disagree	38.11	36.51	39.69
Disagree	38.30	38.40	38.21
Agree	17.71	18.86	16.58
Strongly agree	5.54	5.96	5.12
Missing	0.34	0.26	0.41
I learn maths easily			
Strongly disagree	7.56	8.74	6.40
Disagree	19.89	21.97	17.84
Agree	44.70	45.61	43.81
Strongly agree	27.50	23.39	31.56
Missing	0.35	0.30	0.39
I have fun doing maths			
Strongly disagree	20.62	22.13	19.13
Disagree	22.87	24.46	21.30
Agree	31.95	31.76	32.13
Strongly agree	24.29	21.43	27.11
Missing	0.27	0.22	0.33
I'd like to do more maths at school			
Strongly disagree	37.12	39.63	34.63
Disagree	28.82	29.55	28.10
Agree	20.68	19.72	21.63
Strongly agree	13.18	10.92	15.42
Missing	0.20	0.18	0.22
I believe that being good at Maths will help me in life	**Year 10 (%)**		
Strongly disagree	6.14	5.64	6.62
Disagree	27.4	29.43	25.59
Agree	51.49	51.98	50.98
Strongly agree	14.33	12.35	16.21
Missing	0.60	0.60	0.60
I need to understand Maths in order to learn other subjects at school			
Strongly disagree	10.68	12.28	9.16
Disagree	35.48	39.71	31.46
Agree	42.01	39.07	44.80
Strongly agree	11.20	8.35	13.91
Missing	0.63	0.60	0.67
I need to be good at Maths in order to choose what to do after school			
Strongly disagree	18.65	21.69	15.76
Disagree	34.77	38.19	31.56
Agree	33.41	29.70	36.96
Strongly agree	12.49	9.80	14.97
Missing	0.69	0.62	0.75
I need to be good at Maths in order to get a good job			
Strongly disagree	18.78	21.85	15.85
Disagree	31.87	34.50	29.42
Agree	32.93	30.61	35.17
Strongly agree	15.71	12.37	18.80
Missing	0.72	0.68	0.76

Table A3 – Factor Analysis. Attitudes towards maths

Factor	Eigenvalues	Variables
Year 5		
Math self-concept	0.7635	I am good at maths
	0.8133	Maths is hard
	0.7943	I learn maths easily
	0.2617	I have fun doing maths
	0.0754	I'd like to do more maths a school
Year 6		
Math self-concept	0.7737	I am good at maths
	0.6509	Maths is hard
	0.7997	I learn maths easily
	0.7864	I have fun doing maths
	0.7146	I'd like to do more maths a school
Year 10		
Importance of math for the future	0.7054	I believe that being good at Maths will help me in life
	0.7429	I need to understand Maths in order to learn other subjects at school
	0.7887	I need to be good at Maths in order to choose what to do after school
	0.7907	I need to be good at Maths in order to get a good job

Table 3 – Effect of other independent variables on achievements in Mathematics (OLS and School FE)

	OLS					**School Fixed effects**				
	Year 2	**Year 5**	**Year 6**	**Year 8**	**Year10**	**Year 2**	**Year 5**	**Year 6**	**Year 8**	**Year10**
Escs index	n.a	0.109	0.172	n.a	0.151	n.a	0.098	0.168	n.a	0.035
		(0.012)***	(0.012)***		(0.009)***		(0.012)***	(0.013)***		(0.008)***
Escs*Female	n.a.	0.005	-0.017	n.a.	-0.029	n.a.	0.0065	-0.025	n.a.	-0.008
		(0.013)	(0.012)		(0.010)***		(0.013)	(0.013)*		(0.008)
Region of residence										
(North west is omitted)										
North-East	0.057	-0.013	-0.087	0.054	0.016	-	-	-	-	-
	(0.022)***	(0.022)	(0.021)***	(0.021)***	(0.016)					
Centre	-0.047	-0.157	-0.276	-0.119	-0.376	-	-	-	-	-
	(0.023)**	(0.023)***	(0.021)***	(0.023)***	(0.016)***					
South	-0.210	-0.264	-0.432	-0.227	-0.630	-	-	-	-	-
	(0.021)***	(0.020)***	(0.018)***	(0.021)***	(0.015)***					
Islands	-0.247	-0.405	-0.654	-0.055	-0.773	-	-	-	-	-
	(0.024)***	(0.023)***	(0.021)***	(0.022)**	(0.016)***					
Maternal education										
(University is omitted)										
High school	-0.118	-0.140	-0.037	-0.159	-0.028	-0.115	-0.129	-0.045	-0.122	0.001
	(0.023)***	(0.023)***	(0.022)	(0.023)***	(0.016)**	(0.022)***	(0.022)***	(0.022)**	(0.022)***	(0.013)
Middle school	-0.316	-0.318	-0.237	-0.378	-0.151	-0.299	-0.315	-0.237	-0.357	-0.025
	(0.027)***	(0.028)***	(0.027)***	(0.025)***	(0.019)***	(0.026)***	(0.027)***	(0.026)***	(0.024)***	(0.015)*
Paternal education										
(University is omitted)										
High school	-0.106	-0.022	-0.036	0.018	-0.000	-0.109	-0.044	-0.058	0.016	0.039
	(0.023)***	(0.026)	(0.024)	(0.022)	(0.016)	(0.023)***	(0.024)*	(0.024)**	(0.021)	(0.013)***
Middle school	-0.316	-0.170	-0.162	-0.378	-0.125	-0.310	-0.202	-0.201	-0.196	0.005
	(0.027)***	(0.020)***	(0.027)***	(0.025)***	(0.019)***	(0.024)***	(0.028)***	(0.027)***	(0.022)***	(0.015)
N. siblings	n.a.			n.a.		n.a			n.a	
(0 is omitted)										
1		-0.017	0.016		0.044		-0.004	0.009		0.048
		(0.020)	(0.019)		(0.010)***		(0.019)	(0.019)		(0.012)***
2		-0.066	-0.044		0.005		-0.057	-0.026		0.052
		(0.023)**	(0.022)**		(0.017)		(0.023)**	(0.022)		(0.014)***
3		-0.071	-0.112		0.012		-0.095	-0.105		0.091
		(0.037)*	(0.035)***		(0.027)		(0.036)*	(0.035)***		(0.022)***
>4		-0.183	-0.211		-0.099		-0.217	-0.185		0.030
		(0.051)***	(0.047)***		(0.037)***		(0.049)***	(0.047)***		(0.030)

Note: Std errors are in brackets. * indicates that the underlying coefficient is significant at 10% level, ** at 5% and ***1%.